Effectiveness of New Weibull Analysis Methods of Life Data Including Suspended Ones

K. SHITSUKAWA* M. SHIBATA**

* Bearing Engineering Center, Automotive Unitized Product Engineering Department
** Research & Development Center, Bearing Engineering Development Department

New Weibull analysis methods which can estimate confidence bounds of life data including suspended ones have been developed and available.

Here, calculation algorithm and effectiveness of the Fisher's Matrix Bounds method which is widely used among various analysis softwares are discussed.

***Key Words**: Weibull analysis, confidence bound, maximum likelihood, fisher's matrix*

1. Introduction

In order to assure the reliability of machine parts (e.g. bearings, gears, etc.), conventionally endurance testing has been carried out, and the test data have been analyzed using the Weibull analysis. However, for higher reliability, a lot of test hours and costs have been demanded because the conventional Weibull analysis needs more failure data. Furthermore, for bearings, it has become difficult to obtain the amount of fatigue life data necessary for the analysis, because in recent years, bearing fatigue lives have become significantly longer under cleaner lubricating conditions[1].

For the purpose of shortening testing times, the analysis method had been already established by L. G. Johnson which estimates the life value more precisely also by the test data including suspended ones[2]. But the calculation method of confidence bounds when the life data include suspended ones had not been mentioned. Therefore, in Japan it has not been tried to calculate the confidence bounds, and the documents which mentions the calculation method have not been published. On the other hand, the demand for higher reliability machine parts by the users has been increasing all the more. And the data analysis on higher reliability than conventional 90% has been demanded.

Under those background, R. B. Abernethy et al have developed the analysis software named Super SMITH by which the confidence bound for life data including suspended ones can be calculated on higher reliability[3] (In this paper the methods are called Abernethy's methods). However, the validity and detail of the analysis have not been explained very clearly in the manual of Super SMITH. Therefore, the authors have been studying the mathematical strictness and verifying the validity of the analysis methods programmed in Super SMITH.

This paper describes the verification result by the authors about the calculation methods of following two confidence bounds.

1. Beta-Binomial Bounds
   This method had been basically established by L. G. Johnson and after that, has been improved to handle with the life data including suspended ones in the United States.
2. Fisher's Matrix Bounds
   It is thought this method is original and popular in the United States but it is scarcely known in Japan.

2. Conventional Calculation method of Confidence Bounds (Beta-Binomial Bounds) and the Problems

2. 1 Calculation Method

The basic Weibull analysis is mentioned very little in this paper because some excellent references have been published[4)~7)].

The conventional calculation method of confidence bounds had been established by L. G. Johnson[2] and is called Beta-Binomial Bounds[3].

The calculation method is given by Eqs. (1)~(3).

$$t_{j,L} = \eta \left\{ \ln \left(\frac{1}{1-F_{j\{(100-C)/200\}}} \right) \right\}^{1/\beta} \tag{1}$$

$$t_{j,U} = \eta \left\{ \ln \left(\frac{1}{1-F_{j\{(100+C)/200\}}} \right) \right\}^{1/\beta} \tag{2}$$

$$\begin{aligned} &{}_nC_j\lambda^j(1-\lambda)^{n-j} + {}_nC_{j+1}\lambda^{j+1}(1-\lambda)^{n-j-1} + \cdots \\ &\cdots + {}_nC_{n-1}\lambda^{n-1}(1-\lambda) + {}_nC_n\lambda^n = A \end{aligned} \tag{3}$$

The left side of the equation (3) is the cumulative distribution function of an order statistics and the solution λ is the Median Ranks when the right hand A of the equation is equal to 0.5.

The confidence bounds can be calculated by solving the Eq. (3) and substituting λ for $F_{i\{(100+C)/200\}}$ and $F_{i\{(100-C)/200\}}$ in the Eqs. (1) and (2) each other.

However, when life data include suspended ones, it is very difficult to solve the Eq. (3) because an order number j is not always an integer.

The authors have researched the solution. As a result, it has been found that it is not impossible to solve Eq. (3) when *j* is not an integer, but that the solution is clearly irrational.

Abernethy's method states that each rank for the order number which is not an integer for the existence of suspended data is calculated by linear interpolation from each rank supposing that all data are failure data (when an order number is an integer), and is substituted for Eqs. (1) and (2)[8), 9)]. The method is explained using the following example based upon the life testing data of deep groove ball bearings which is carried out under the condition of clean oil.

The example of life testing data including suspended ones is shown in **Table 1**. For the sake of simplicity, the confidence bounds under C = 90% are calculated.

Table 1 Assignment of all fatigue life data to the each rank

Order number *j*	Life time, h	Median ranks	5%ranks	95%ranks
1	125	0.1294	0.0102	0.4507
2	(238)	0.3138	0.0764	0.6574
3	339	0.5000	0.1893	0.8107
4	503	0.6862	0.3426	0.9236
5	846	0.8706	0.5493	0.9898

In Abernethy's method, to begin with, the order number and each rank are assigned to all life data including suspended ones as shown in **Table 1**. As each rank is written in various references (for example (3)[3)]) in this case(C = 90%), Eq. (3) need not be solved. Followingly, as shown in **Table 2**, the mean order numbers are calculated and they are assigned to only failure data. And each new rank which corresponds to each mean order number is calculated by linear interpolation using the order number and each rank in **Table 1**. For example, new 5% ranks which corresponds to j' = 2.25 is calculated in the following way.

$$0.0764 + (0.1893–0.0764)\ (2.25–2)/(3–2) = 0.1046$$

Finally, the upper and lower confidence bounds are calculated by substituting the new 5% ranks and 95% rank in **Table 2** into Eqs. (1) and (2). The calculated confidence bounds are shown in **Fig. 1**. Here, the bearing life (in hours) at the new median ranks are designated by △ and the lives at new 5% and 95% ranks are shown as ○.

Table 2 New ranked data

Mean order number *j'*	Life time, h	Median ranks	5%ranks	95%ranks
1	125	0.1294	0.0102	0.4507
	(238)			
2.25	339	0.3604	0.1046	0.6957
3.5	503	0.5931	0.2660	0.8672
4.75	846	0.8245	0.4976	0.9733

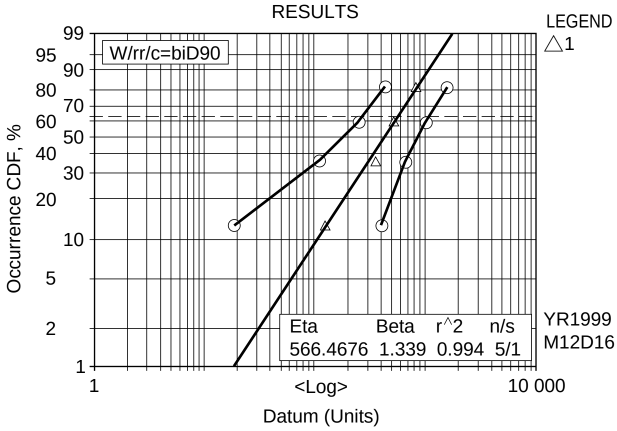

Fig. 1 Bata-Binominal 90% confidence bounds

The new median ranks are calculated in the same method also in the conventional Weibull analysis and it is recognized all over the world. 5% ranks and 95% ranks are the same order statistics as Median ranks. Therefore, Abernethy's methods can be regarded as the naturally developed one which copes with them in the same way, and it is considered that it does not include theoretical contradiction.

2. 2 Limit of Beta-Binomial Bounds

Conventional confidence bounds are represented by points equal to number of failure data as shown in **Fig. 1**. Generally, a curve is drawn between the points. But the curve can not be drawn over the top point and under the bottom point, because the change is not known at all in the region. Therefore, the confidence bounds at the reliability over the top point and under the bottom point can not be estimated. Moreover, the estimated region becomes narrow as failure data become few. For example, as for **Fig. 1**, the lower confidence bounds at L_{10} life can not be estimated. It is considered that lower confidence bounds are needed to estimate expecting safety in order to assure confidence , especially when failure data are few, accordingly low confidence. But the present method can not solve this problem.

2. 3 Outline of New Confidence Bounds Calculation Methods

After 1980, new confidence bounds calculation methods have been suggested one after another instead of Beta-Binomial Bounds in the United States. The characteristics of them are shown in **Table 3**.

As shown in the same Table, four kinds of calculation methods are suggested. And they can be calculated by the Super SMITH software together with the conventional method(Beta-Binomial Bounds). Calculation results are somewhat different from each other (fairly different depending upon the conditions). The four kinds of calculation methods are improved to solve the problem in the conventional one, and have the following features in common.

1) Although calculation results of the conventional method have discontinuous points, calculation results of the four kinds of new methods are continuous curves.

2) Although the conventional method cannot calculate in low (or high) cumulative failure probability region, the four kinds new methods can estimate in all region (0% < cumulative failure probability < 100%).

Table 3 New calculation methods of confidence bounds

Name	Point
Fisher's Matrix Bounds[10), 11)]	The nature that the distribution of the maximum likelihood estimates follow the normal distribution is used. Local Fisher's information matrix is used for the purpose of calculation of variance.
Likelihood Ratio Bounds[10)]	The nature that likelihood ratio statistic Λ is submitted to the chi-square distribution with the degree of freedom 1 is used.
Monte Carlo Simulation Bounds[12)]	Monte Carlo Simulation (the method that random number is used) is used.
Pivotal Bounds[10)]	Each % point of the distribution of the Pivotal Quantity is used.

By the way, it seems that among the four kinds of new methods Fisher's Matrix Bounds are used the most popularly in the United States and the method is default in the Super SMITH software. Therefore, only the method will be explained theoretically bellow. First of all, the maximum likelihood estimation will be investigated to compare with least squares method which is used popularly because the new methods use maximum likelihood estimation as recurring method.

2. 4 Maximum Likelihood Estimation[13), 14)]

The comparison between Weibull analysis with maximum likelihood estimation and that with least squares method is shown in **Table 4**.

$$\ln L = \ln\left[\prod_{i=1}^{r} f(x_i)\prod_{j=1}^{k}\{1-F(T_j)\}\right] \quad (4)$$

$$\ln L(\beta,\eta) = r\ln\beta - r\beta\ln\eta + (\beta-1)\sum_{i=1}^{r}\ln x_i - \sum_{i=1}^{n}(t_i/\eta)^{\beta} \quad (5)$$

$$\varepsilon = X_n - (Y_n - a)/\beta \quad (6)$$

$$Q_n = \Sigma\,\varepsilon^2 \quad (7)$$

$$\frac{\partial}{\partial\beta}(\ln L) = 0 \quad (8)$$

$$\frac{\partial}{\partial\eta}(\ln L) = 0 \quad (9)$$

The estimates by maximum likelihood estimation are usually symbolized by $\hat{\beta}$, $\hat{\eta}$ etc. and named maximum likelihood estimates.

Table 4 Comparison of maximum likelihood and least squares method

	Maximum likelihood estimation	Least squares method
Estimate	Weibull slope β characteristic life η (Constants in population is estimated.)	Weibull slope β y cut of recurring line : **a** (Constants in sample is estimated.)
Function	Logarithmic likelihood function $\ln L$ (β, η) Eqs. (4) and (5)	Square sum of tolerance Q_n (β, **a**) : **Fig. 3**, Eqs. (6) and (7)
Mathematical condition	$\ln L$ is maximum. (The coordinates of the top of the curved surface in **Fig. 2** is estimate) $\frac{\partial}{\partial\beta}(\ln L)=0$ $\frac{\partial}{\partial\eta}(\ln L)=0$	Q_n is minimum. $\frac{\partial Q_n}{\partial\beta}=0$ $\frac{\partial Q_n}{\partial a}=0$
Reliability	The most excellent recurring method when number of data is many ($n \geqq 25$), the larger Weibull slope is estimated when number of data is not many ($n \leqq 10$)	More excellent than likelihood estimation when number of data is not many ($n \leqq 10$)

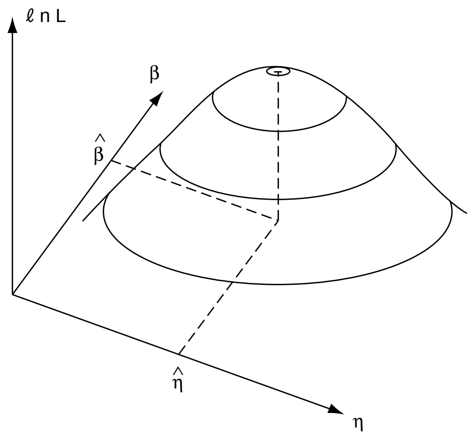

Fig. 2 The logarithmic likelihood function

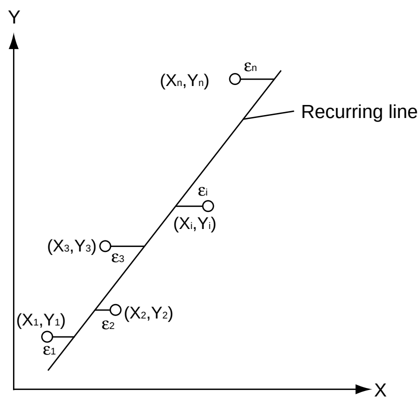

Fig. 3 The method of least squares

In order to calculate the estimates concretely Eqs. (10) and (11) which are obtained by substituting Eq. (5) for Eqs. (8) and (9) and by arranging them are solved numerically.

$$\frac{\sum_{i=1}^{n} t_i^{\hat{\beta}} \ln t_i}{\sum_{i=1}^{n} t_i^{\hat{\beta}}} - \frac{1}{r}\sum_{i=1}^{r} \ln x_i - \frac{1}{\hat{\beta}} = 0 \tag{10}$$

$$\hat{\eta} = \left[\frac{1}{r}\sum_{i=1}^{n} t_i^{\hat{\beta}}\right]^{1/\hat{\beta}} \tag{11}$$

2. 5 Theoretical Formulae of Fisher's Matrix Bounds

The data which have changed t_i following the Weibull distribution by Eq. (12) follow the distribution named the extreme value distribution. The extreme value distribution is introduced in order to simplify the theoretical formulae.

$$y_i = \ln t_i \tag{12}$$

In case of large sample (Number of data is numerous), the maximum likelihood estimates $\hat{y}_p$ of p% point y_p of the extreme value distribution follows the normal distribution which has mean value y_0, asymptotic variance Asvar (y_p). The reason is that maximum likelihood estimates are estimated aiming at the true value y_p and as a result they follow the normal distribution which has it as a mean value. Therefore, $(\hat{y}_p - y_p)/\{As\ \mathrm{var}\ (\hat{y}_p)\}^{1/2}$ follows the standard normal distribution, the y_p' upper confidence bounds $y_{p,U}$ are given by Eq. (13) and the y_p' lower confidence bounds $y_{p,L}$ are given by Eq. (14).

$$y_{p,U} = \hat{y}_p + U(P)\,\{As\ \mathrm{var}\ (\hat{y}_p)\}^{1/2} \tag{13}$$

$$y_{p,L} = \hat{y}_p - U(P)\,\{As\ \mathrm{var}\ (\hat{y}_p)\}^{1/2} \tag{14}$$

The y_p is given by Eqs. (15) and (16) using location parameter u and scale parameter b of the extreme value distribution.

$$y_p = u + w_p b \tag{15}$$

$$w_p = \ln\{-\ln(1 - p/100)\} \tag{16}$$

And Eq. (17) is induced by the nature of maximum likelihood estimates.

$$\hat{y}_p = \hat{u} + w_p \hat{b} \tag{17}$$

The calculation method of $\hat{u}$ and $\hat{b}$ is omitted. They are obtained by Eqs. (4), (8) in the same way as $\hat{\beta}$, $\hat{\eta}$ of the Weibull distribution.

Secondly, as for asymptotic variance As var $(\hat{y}_p)$, Eq. (18) is induced.

$$As\ \mathrm{var}\ (\hat{y}_p) = As\ \mathrm{var}\ (\hat{u}) + 2w_p\, As\ \mathrm{cov}\ (\hat{u}, \hat{b}) + w_p^{\,2}\, As\ \mathrm{var}\ (\hat{b}) \tag{18}$$

By the way, the matrix expression of each term of right part of Eq. (18) is called the covariance matrix. And it is equal to the inverse matrix of the matrix I_0 called local Fisher's information matrix. The matrix is expressed as Eqs. (19) and (20).

$$\begin{bmatrix} As\ \mathrm{var}\ (\hat{u}) & As\ \mathrm{cov}\ (\hat{u}, \hat{b}) \\ As\ \mathrm{cov}\ (\hat{u}, \hat{b}) & As\ \mathrm{var}\ (\hat{b}) \end{bmatrix} = I_0^{-1} \tag{19}$$

$$I_0 = \begin{bmatrix} -\partial^2 \ln L/\partial u^2 & -\partial^2 \ln L/\partial u \partial b \\ -\partial^2 \ln L/\partial u \partial b & -\partial^2 \ln L/\partial b^2 \end{bmatrix}_{(\hat{u}, \hat{b})} \tag{20}$$

In case of the extreme value distribution, the local Fisher's information matrix is expressed by Eqs. (21) and (22).

$$I_0 = \frac{1}{\hat{b}^2}\begin{bmatrix} r & \sum_{i=1}^{n} \hat{z}_i \exp \hat{z}_i \\ \sum_{i=1}^{n} \hat{z}_i \exp \hat{z}_i & r + \sum_{i=1}^{n} \hat{z}_i^{2} \exp \hat{z}_i \end{bmatrix} \tag{21}$$

$$\hat{z}_i = (y_i - \hat{u})/\hat{b} \tag{22}$$

Therefore, when the local Fisher's information matrix is obtained by the same Equations, and the inverse matrix is calculated, the asymptotic variances As var $(\hat{y}_p)$ is obtained by Eqs. (18) and (19), and the confidence bounds of p% point of the extreme value distribution is calculated.

Finally, when the confidence bounds of the extreme value distribution is transformed to the ones of the Weibull distribution by Eq. (23), the confidence bounds of p% life L_p of the Weibull distribution is obtained.

$$L_p = \exp(y_p) \tag{23}$$

By the way, it should be noted that the p of p% life L_p in Eqs. (13), (14), and (23) is not the constants but the symbol. Therefore, it is found that the confidence bounds given by this method are expressed by continuous curves clearly and that cumulative failure probability p can be obtained the all numerical value in the region of $0 < p < 100$.

This nature in this method is quite different from the one in the conventional method (Beta-Binomial Bounds), that is discontinuous points assemble, cumulative failure probability p can be given only the numerical value between the shortest life point and the longest life point.

2. 6 Problems in New Confidence Bounds Calculation Methods

In this paper Fisher's Matrix Bounds have been explained as one of the new confidence bounds calculation methods (Abernethy's method). This method is very significant in practical use on the point that the confidence bounds in the region of high reliability can be obtained. However, the authors have found the following problems:

1) Theoretically, the Abernethy's method is based upon the maximum likelihood estimation. But as shown in **Table 4**, when number of failure data is low (r ≦10), the Weibull slope β is estimated larger by using the maximum likelihood estimation.
2) Fisher's Matrix Bounds are constructed on the assumption of the large sample (failure data are numerous). It is doubtful that they are used when failure data are limited (r ≦10).
3) In the practical calculation example which is stated in the manual of Super SMITH software[3], Fisher's Matrix

Bounds for the ten failure data are calculated not by the maximum likelihood estimation but by the least square method. It is considered that the least square method is adopted because the recurring precision of the least square method is higher than that of the maximum likelihood estimation when failure data is not many. Because theoretically the Fisher's Matrix Bounds are based upon maximum likelihood estimation, it is considered that this treatment is bold.

The Fisher's Matrix Bounds of the bearing life data shown in the foregoing **Table 1** are shown in **Fig. 4 (a)** and **4 (b)** which are calculated by the authors using the maximum likelihood estimation and the least square method respectively. In this case, because the number of failure data is 4, that is, small, the Weibull slope 1.9 calculated by maximum likelihood estimation is larger than the one 1.3 calculated by the least square method. And the lower 90% confidence bounds 63 (hrs) at the L_{10} life estimated by the maximum likelihood estimation is longer than the one 28 (hrs) by the least square method.

In this way, in case of few failure data, there is a problem for the validity of the application of Fisher's Matrix Bounds. The problem of the proper data number shown in this paper needs to be solved in the future because an analysis with higher precision by fewer data is needed in practical use.

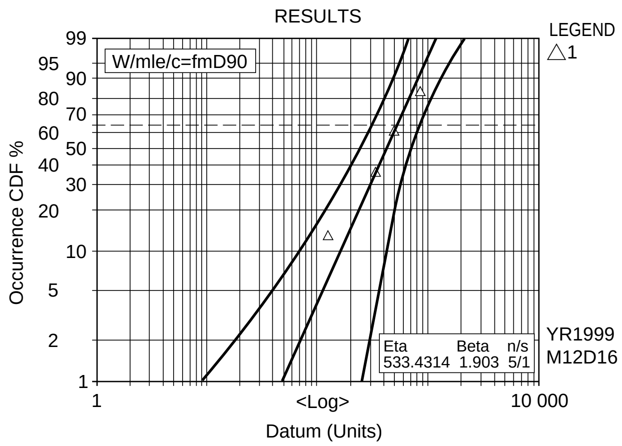

Fig. 4 (a) The Fisher's matrix 90% confidence bounds by the maximum likelihood method

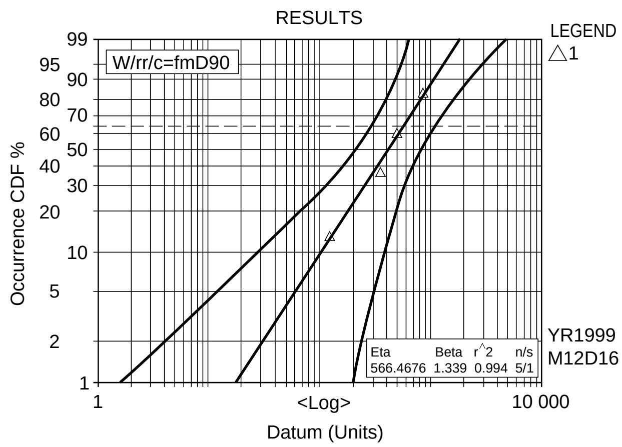

Fig. 4 (b) The Fisher's matrix 90% confidence bounds by the method of least squares

3. Conclusions

About the new calculation method of the confidence bounds in Weibull analysis in case of the data including suspended ones, the results of the study for Fisher's Matrix Bounds among the conventional method and the recent developed ones are shown in the following.

1) Although the conventional method is a natural extended one from the Johnson's method, the estimates of confidence bounds on the reliability over the range which the failure data is plotted in a Weibull probability paper is impossible.
2) Fisher's Matrix Bounds are estimated using the nature which the distribution of the maximum likelihood estimates follows the normal distribution, and using the local Fisher' information matrix for the purpose of calculation of the variance. This method is very useful because the confidence bounds are calculated as a continuous curve in all range of reliability.
3) However, the least square method is used as the recurring method in case of few failure data. This is contradictory to the theory, because the Fisher's Matrix Bounds are based on the maximum likelihood estimation as the recurring method. And the problem of the proper data number needs to be solved in the future.

<Nomenclature>

a	: y cut of recurring line
A	: Constants in Eq. (3) $(100-C)/200$ when $(100-C)/2$%ranks are obtained $(100+C)200$ when $(100+C)/2$%ranks are obtained
As cov	: Asymptotic covariance (covariance when number of data is numerous)
As var	: Asymptotic variance (variance when number of data is numerous)
b	: Scale parameter of the extreme value distribution
$\hat{b}$	: Maximum likelihood estimate of b
C	: Probability of confidence (%)
${}_nC_j$	: Binomial coefficient ($= n!/j!(n-j)!$)
exp	: Exponential function
$f(x)$	: Probability density function
$F(t)$	: Cumulative failure probability
$F_{j\{(100-C)/200\}}$	: $(100-C)/2$% ranks of life data at number j
$F_{j\{(100+C)/200\}}$	: $(100+C)/2$% ranks of life data at number j
I_0	: Local Fisher's information matrix
I_0^{-1}	: Inverse matrix of local Fisher's information matrix
j	: Order number of life data
j'	: Mean order number
k	: Number of suspended life data ($r + k = n$: number of all life data)
ln	: Napierian logarithm
ln$L$	: Logarithmic likelihood function

$\ln L(\beta, \eta)$: Logarithmic likelihood function of the Weibull distribution with two constants in the population
L_p : p% life of the Weibull distribution
n : Number of all life data including suspended ones
p : Each p% point
P : Both sides probability of the standard normal distribution (=1 − C/100)
Q_n : Square sum of tolerance
r : Number of failure data
R : Reliability (1 – $F(t)$)
u : Location parameter of the extreme value distribution
$\hat{u}$: Maximum likelihood estimate of u
$U(P)$: One side 100P/2% point of the standard normal distribution
t_i : All life data including failure data and suspended ones
T_j : Suspended life data
$t_{i,L}$: C% lower confidence bounds of life data at number i
$t_{j,U}$: C% upper confidence bounds of life data at number j
w_p : The value defined by Eq. (16)
x_i : Failure data
X_n : X coordinates of data point on the Weibull probability paper
y_i : The dada submitted to the extreme value distribution
Y_n : Y coordinates of data point on the Weibull probability paper
y_P : p% point of the extreme value distribution
$\hat{y}_p$: Maximum likelihood estimate of y_p
$y_{p,U}$: Upper confidence bounds of y_p
$y_{p,L}$: Lower confidence bounds of y_p
$\hat{z}_i$: The value defined by Eq. (24)
β : Weibull shape parameter or Weibull slope
$\hat{\beta}$: Maximum likelihood estimate of β
η : Weibull scale parameter or characteristic life
$\hat{\eta}$: Maximum likelihood estimate of η
λ : Value of each ranks of life data number j (solution of Eq.3)
ε : Tolerance
$\prod$: All product

References

1) K. Shitsukawa, M. Shibata, Y. Ohno and T. M. Johns: SAE Paper no. 972710 (1997).
2) L. G. Johnson: The Statistical Treatment of Fatigue Experiments, Elsevier Publishing Company (1964).
3) R. B. Abernethy: The New Weibull Handbook 2nd edition, (1994) distributed by SAE.
4) Y. Tateishi: Koyo Engineering Journal, No. 130 (1986) 47.
5) M. Satou, J. Okamoto: Bearing, **28**, 5 (1985) 2.
6) M. Satou, J. Okamoto: Bearing, **28**, 6 (1985) 4.
7) M. Satou, Tribologyst, **39**, 8 (1994) 49.
8) L. Charles, et al: Statistical Design and Analysis of Engineering Experiments, McGraw-Hill, (1973).
9) K. C. Kapur and L. R. Lamberson: Reliability in Engineering Design, John Wiley, (1977).
10) J. F. Lawless: Statistical Models & Methods for Lifetime Data, John Wiley, (1982).
11) W. Nelson: Applied Life Data Analysis, John Wiley (1982).
12) W. Q. Meeker, et al: Statistical Methods for Reliability, John Wiley (1998).
13) T. Kawada, K. Kunisawa: Modern Statistics first volume revision, Hirokawashoten (1967).
14) T. Ichida, K. Suzuki,: Distribution and Statistics for Reliability, Nikkagirenshuppansha (1984).